PRIMARY NUCLEOTIDE SEQUENCE OF THE SHRIMP WHITE SPOT BACILLIFORM VIRUS (WSBV), DISCOVERY SYSTEMS CONTAINING THIS SEQUENCE AND DETECTION KITS AND ANTIVIRAL TARGETS FOR DETECTION AND CONTROLLING SHRIMP VIRUS OUTBREAK AND SPREAD

RELATED APPLICATION

The present application claims priority to Chinese patent application No. 99124717.5, filed November 24, 1999.

FIELD OF THE INVENTION

The present invention is in the field of genomic discovery systems. The present invention specifically provides the complete shrimp white spot bacilliform virus (WSBV) genome and isolated fragments thereof in a form that is commercially useful, including detection kits, antiviral agents, reagents such as nucleic acid arrays, and computer-based systems.

BACKGROUND OF THE INVENTION

The shrimp and prawn (hereafter collectively referred to as shrimp) industry is a rapid growth worldwide industry worth billions of dollars. Worldwide, the shrimp industry relies on both the harvesting of wild shrimp and aquaculture, which is the controlled farming of fish, shellfish, and plants. Aquaculture, particularly aquaculture of shrimp, is growing rapidly due to increasing consumer demand for shrimp and other seafood. Aquaculture has been expanding at an annual rate far surpassing the growth of livestock meat, capture fisheries, and agricultural production. The aquaculture industry delivers high-quality protein for human and animal consumption and provides a substantial source of income and employment, particularly for developing countries. Aquaculture accounts for nearly 20 percent of the world's harvest of fish, shellfish, and seaweeds. The total worldwide value of giant tiger prawn production is the greatest of any aquaculture species. Aquaculture of giant tiger prawn significantly contributes to many Asian and Latin American economies, where the majority of giant tiger prawn production occurs. Shrimp accounted for approximately a quarter of the overall value of Asian fish exports in 1996. In the United States, harvesting and processing shrimp, including both aquaculture and harvesting wild shrimp, is a $3 billion dollar a year industry that employs over 11,000 people. Furthermore, shrimp aquaculture in the U.S. has

the potential to become a high-growth business. The risk of viral diseases to cultured shrimp is the primary obstacle to the growth of the shrimp aquaculture industry.

Shrimp viral disease is a major worldwide concern of the shrimp industry. Both aquaculture and wild shrimp are vulnerable to viral infection, which can lead to devastating economic consequences. Furthermore, shrimp viruses may affect other crustaceans such as crabs and crayfish. Drastic declines in the populations of wild shrimp or other crustaceans due to viral disease can also dramatically affect other species in the food chain that depend upon shrimp for food and can lead to severe ecological consequences.

Major pathogenic shrimp viruses include White Spot Bacilliform Virus (WSBV), Infectious Hypodermal and Hematopoietic Virus (IHHNV), Tara Syndrome Virus (TSV), and Yellow Head Virus (YHV). IHHNV and TSV are endemic throughout South and Central America, while WSBV and YHV are endemic throughout Asia. All U.S. shrimp species are susceptible to infection and disease from one or more of these four viruses. Susceptibility of U.S. species of shrimp to these viruses may lead to restrictions on the importation of foreign shrimp into the U.S.

Past incidents of viral outbreaks illustrate the devastating affects that a viral outbreak can have on the shrimp industry. An outbreak of IHHNV in 1987 in the Gulf of California shrimp fishery reduced shrimp to levels that could not support commercial harvests until 1994. Outbreaks in 1995 and 1996 on U.S. shrimp farms caused a 50 to 95 percent loss of production at affected farms. Shrimp exports from China to the U.S. dropped 75% between 1990 and 1995 due to infection by WSBV.

WSBV is regarded as one of the most highly pathogenic viruses of penaeid shrimp. No uniform name exists for WSBV. It is also known as White Spot Syndrome Virus (WSSV), Prawn White Spot Bacilliform Virus (PWSBV), White Spot Baculiform Virus (WSBV), Baculoviral Hypodermal and Hematopoietic Necrosis Virus (HHNBV), Rod-shaped Nuclear Virus of Penaeus japonicus (RV-PJ), Systemic Ectodermal and Mesodermal Bacilliform Virus (SEMBV). Other acronyms include WSV, WSDV, and LNBV. The virus is a non-occluded, circular, double-stranded DNA bacilliform virus with a genome of approximately 300kb. WSBV virions are enveloped nucleocapsids with bacilliform morphology and a tail-like extension at one end.

White Spot Syndrome, caused by WSBV, is also known by such names as Red Disease, China Virus Disease, and Shrimp Explosive Epidemic Disease. Infected shrimp display rapid reduction in food consumption and lethargy. Gross observations include a loose cuticle and a red color to the entire body and appendages along with small subcutaneous

white spots. Histological examination reveals prominent intranuclear inclusion bodies in the cuticular epithelium, subcutis, and connective tissues. Cumulative mortality rates reach 100% within 3 to 10 days of the onset of clinical signs. No significant resistance to WSBV has been reported. All native U.S. species of shrimp are susceptible to WSBV infection under experimental conditions. WSBV is widely spread throughout most of the shrimp growing regions of Asia and the Indo-Pacific, including China, Japan, Korea, Thailand, Indonesia, Taiwan, Vietnam, Malaysia, and India. Lethal outbreaks of WSBV virus have recently been recorded in Texas and South Carolina. Furthermore, the virus has been shown to infect other crustaceans including amphipods, ostracods, swimming crabs, crayfish, copepods, and shore flies. The possibility exists that these organisms could act as a reservoir through which further shrimp infection, or infection of other species, can occur.

In view of the serious economic and ecological risks posed to the worldwide shrimp industry and shrimp populations by viruses, particularly WSBV, a strong need exists for antiviral agents and detection systems. Detection systems should be highly specific, rapid, and sensitive. To facilitate development of antiviral agents and detection systems, knowledge of the complete genomic sequence and protein encoding sequences of WSBV is needed. Prior to the present invention, very few reports on WSBV genomic sequences existed, and only a small fraction of the entire WSBV genome had been sequenced. To date, only six WSBV sequenced have been patented, published or stored in public genome databases, such as Genbank. All sequences to date are short sequences ranging in length from 420 bp to 2424 bp. J.S. Kim and others from Korea have sequenced 2424bp (wsu 92007,1997) and 420bp (wsu 89843, 1997); K.Mitsuo and others from Japan have sequenced two fragments, 1447bp (PN JP 1997201196-A/2) and 1461bp (PN JP 1997201196-A/1) in length; Chufang Luo et.al. from Taiwan sequenced 1461bp (PMU50923, 1996); L.M. Nunan et. al. from the United States has reported 868bp sequence in *J. Virological Methods* (1997(63): p193-201). These known sequences are no more than 10kb in length all together. In addition, these sequences are randomly sequenced with no systematic analysis; therefore determining sequence function is difficult. Since the complete genome of WSBV is more than 300kb in length, the analysis of the complete genomic DNA sequence and it's complete structure, the determination of the expressed sequences, and prediction of the functions of encoded proteins are all new scientific achievements. These achievements are the basis for the present invention. The present invention is directed to providing the complete primary nucleotide sequence of WSBV and isolated fragments thereof, protein encoding sequences of WSBV,

and antiviral agents and detection systems based on the nucleotide and protein encoding sequences provided by the present invention.

DNA Viruses

Generally, transcription of a DNA virus genome occurs in the nucleus of the host cell, utilizing host cell polymerases and other host enzymes for viral mRNA synthesis and viral replication. Viral gene transcription is modulated by the interaction of specific DNA-binding proteins with promoter and enhancer elements in the viral genome. Commonly, the viral promoter and enhancer elements are similar in sequence to those of the host cell in order to allow the host cell's transcriptional activation factors and DNA-dependent RNA polymerase to bind the viral control elements. Cells from different tissues or species express different DNA-binding proteins, and this is a major factor in determining which species, and which cells and tissues of that species, that the virus can infect.

Viruses, in general, depend on the host cell ribosomes, transfer RNA (tRNA), and mechanisms of posttranslational modification to produce their proteins. Generally, viral mRNA encoding non-structural viral proteins, such as DNA-binding proteins and enzymes, are transcribed first. These are followed by late viral gene products encoding structural proteins.

Viruses utilize various methods to promote preferential translation of their viral mRNA over host cell mRNA. In some instances, concentration of viral mRNA in the host cell is so large that it occupies most of the cell's ribosomes, thereby preventing translation of host cell mRNA. Viruses may inhibit synthesis and/or induce degradation of the cell's nucleic acids. Many viruses increase the permeability of the host cell membrane, thereby reducing the ribosomal affinity for most cellular mRNA.

Viral DNA replication begins at a unique sequence in the genome called the origin of replication, or *ori*. The *ori* is recognized by viral or host nuclear factors and DNA-dependent DNA polymerase. Viral DNA synthesis is semi-conservative and a primer is required by the DNA polymerase to initiate synthesis of the new DNA molecule.

Viral Screening Tests and Antiviral Agents

Viral screening tests and detection kits, such as nucleic acid arrays, can be developed based on either nucleic acids or polypeptides provided by the present invention. A nucleic acid probe to a virus specific nucleotide sequence, or an antibody to a virus specific protein,

is introduced into contact with a sample, such as a sample of shrimp cells, whereby the presence of the virus is detected using an assay system.

Antiviral agents, either nucleic acid or protein-based, directly interfere with viral function or preferably, interfere with viral replication to stop or prevent spread of the virus in a population, such as in a population of shrimp. Knowledge of the nucleic acid and protein sequences of the virus allows antiviral agents to be designed to attack a number of viral targets necessary for viral replication or function, such as viral encoded enzymes or structural proteins. Attachment of the virus to the host cell is the first step in viral replication and is mediated by the interaction of a viral attachment protein and a host cell surface receptor. This interaction can be blocked by neutralizing antibodies, which bind to and coat the virion, or receptor antagonists which are peptide or carbohydrate analogues of the viral attachment protein and competitively block the interaction of the virus with the cell. Agents can be designed that bind to the viral attachment protein and prevent penetration of the virus into the cytoplasm or nucleus of the host cell and/or uncoating of the virus. These agents thereby prevent the virus from delivering its genome into the host cell. Viral mRNA expression and utilization can be targeted with anti-viral agents. Antisense oligonucleotides can be designed to bind to newly transcribed viral RNA and thereby prevent the viral RNA from being processed to mRNA in the nucleus, delivered to the cytoplasm, and bound to the ribosome. Many antiviral drugs are nucleoside analogues, which inhibit viral polymerases. Viral polymerases are often less specific for substrate than are host polymerases, therefore the viral polymerase will often bind a nucleotide analogue with a modified base and/or sugar several hundredfold better than the host enzyme. Antiviral drugs can therefore be preferentially incorporated into the viral genome. DNA viruses, such as WSBV, are particularly susceptible to these types of drugs due to the extent and rapid rate of nucleotide incorporation during viral replication. Inhibition of posttranslational modification of viral proteins, such as phosphorylation, may also inhibit viral replication.

SUMMARY OF THE INVENTION

The present invention is based on the sequencing and assembly of the WSBV genome. The present invention provides the primary nucleotide sequence of the WSBV genome (SEQ ID NO: 1) and predicted transcript sequences (SEQ ID NOS: 2, 4, 6...280, 282, 284, 286-293: See the Sequence Listing and the Figure Sheets for both the genomic and transcript sequences) and polymorphic sites on these transcripts summarized in Table 1

hereinafter, and protein encoded sequence produced from each of the genes found in the WSBV genome. This information is provided in the form of sequences and annotation information and can be used to generate computer based discovery systems, nucleic acid detection reagents and kits such as nucleic acid arrays, protein based detection kits, and antiviral targets.

The present invention provides these nucleotide sequences of the WSBV genome, and representative fragments thereof, in a form that can be used, analyzed, and commercialized. For example, the present invention provides the nucleic acid sequences as contiguous strings of primary sequences in a form readable by computers, such as recorded on computer readable media, e.g., magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The present invention specifically provides a Sequence Listing in computer readable form stored on such media. Such compositions are useful in the discovery of drug and antiviral targets.

The present invention further provides systems, particularly computer-based systems that contain the primary sequence information of the present invention stored in data storage means. Such systems are designed to identify commercially important fragments of the WSBV genome.

Another embodiment of the present invention is directed to isolated fragments, and collections of fragments, of the WSBV genome. The fragments of the WSBV genome include, but are not limited to, fragments that encode peptides, hereinafter open reading frames (ORFs) and fragments that modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs). The ORFs are provided in the Sequence Listing and in Figure 3.

The present invention further includes kits, such as nucleic acid arrays, detection reagents and microfluidic devices, that comprise one or more fragments of the WSBV genome of the present invention, particularly ORFs. The kits, such as arrays, can be used to track the expression of many genes, even all genes, or rationally selected subsets thereof, contained in the WSBV genome.

The identification of the entire coding set of sequences from the genome of WSBV will be of great value to all laboratories working with this organism and for a variety of commercial and ecological purposes. Many fragments of the WSBV genome will be immediately identified by similarity searches against protein and nucleic acid databases and

by identifying structural motifs present in protein domains and will be of immediate value to WSBV researchers and for commercial value for controlling WSBV infection in shrimp populations. A specific example concerns viral envelope proteins, many of which interact with host cells. Proteins of this family can readily be configured into screens and assays for detecting chemical modulators of the protein activity. The biological significance of this and other families of proteins for controlling viral replication is well known. Many of the known antiviral agents modulate the activity of these types of proteins. The WSBV genome will allow one to identify all potential antiviral targets.

The present invention is further directed to isolated WSBV proteins encoded by the ORFs of the present invention. A variety of methodologies known in the art can be utilized to obtain any one of the proteins of the present invention. The amino acid sequence can be synthesized using commercially available peptide synthesizers. In an alternative method, the viral protein can be purified from cells infected with the virus.

The invention further provides antibodies that selectively bind one of the WSBV proteins encoded by the present invention. Antibodies have use in viral detection and control and can be generated using the protein encoding sequences provided by the present invention. Such antibodies include both monoclonal and polyclonal antibodies, and fragments thereof. The invention further provides hybridomas capable of producing the above-described monoclonal antibodies.

The present invention provides methods of identifying WSBV in a test sample, such as a sample of shrimp. Such methods comprise incubating cells extracted from the test sample with one or more of the antibodies or probes based on the nucleic acid sequences provided by the present invention under conditions that allow a skilled artisan to determine if the test sample contains the ORF or product produced therefrom.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a protein encoded by one of the ORFs of the present invention. Specifically, such agents include antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise the steps of contacting an agent with an isolated protein encoded by one of the ORFs of the present invention and determining whether the agent binds to said protein.

DESCRIPTION OF THE FIGURE SHEETS

Figure 1 provides a block diagram of a computer system 102 that can be used to implement the computer-based systems of the present invention.

Figure 2 (Sheets 1-40) provides the primary genomic sequence of WSBV.

Figure 3 (Sheets 1-160) provides:

1) the predicted transcript sequence of the WSBV gene and starting ATG site (SEQ ID NOS: 2, 4, 6, 8 . . . 280, 282, 284, 286-293);
2) the predicted protein sequence of the WSBV gene (SEQ ID NOS: 3, 5, 7, 9 . . .281, 283, 285);
3) results of a BLAST query run using default parameters that shows proteins producing significant alignments with the predicted WSBV protein sequence of the present invention.
4) comments
5) TaqMan primer/probe sets. Oligonucleotide sequences useful as primers and/or probes for amplifying and/or screening for the WSBV genes provided by the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

General Description

The present invention is based on the sequencing and assembly of the WSBV genome. In this process, the primary nucleotide sequence of 5795 nucleic acid fragments was determined. These fragments were assembled into a single contiguous sequence of 305,107 bp. After assembly, the sequences were analyzed with various computer packages and compared with all external data sources. The result of this analysis was the identification of 150 predicted genes/transcripts contained in the WSBV genome. The present invention provides the genomic nucleic acid sequence of WSBV (SEQ ID NO: 1), see Figure 2, Sheets 1-40, as well as the predicted gene structure of all 150 identified genes (SEQ ID NOS: 2, 4, 6...280, 282, 284, 286-293) and polymorphic sites on these transcripts summarized in Table 1, and predicted amino acid sequences of all of the encoded proteins (SEQ ID NOS: 3, 5, 7...281, 283, 285), see Figure 3, sheets 1-160.

Table 1. - Summary of SBV Polymorphic Sites

Sites	Polymorphism
13091	T/C
26054	T/G
26074	C/T
48030	T/C
48134	T/C
50395	C/T
76148	A/G
80650	T/A
84076	*/TGC
89028	A/*
93522	T/G
93595	54x10=540bp repetitive sequences/* Within repeats: 89 G/T 143 G/T 197 G/T 251 G/T 305 G/T
97419	CC/**
105400	C/T
108024	A/C
108027	G/*
126375	*/GGAAGAAGAAGAGGAAGA
133906	*/G/GG
139662	G/T
142288	C/T
166492	TGT/***
180850	C/T
181761	TCC/***
200713	G/A
214165	C/T
230870	G/A
232887	G/A
238111	G/A
247636	G/T
258606	***/CTA/CTACTA
271845	C/T
272344	A/C
272751	C/T
273191	G/A
282103	G/*
282189	G/A

The nucleotide sequences of the present invention, or representative fragments thereof, are provided in a form that can be readily used, analyzed, and interpreted by a skilled

artisan. In one embodiment, the sequences are provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences provided in the figures.

As used herein, a "representative fragment of the nucleotide sequence provided herein" refers to any portion of these sequences that are not presently represented within a publicly available database. Preferred representative fragments of the present invention are WSBV open reading frames and expression modulating fragments (ORFs and EMFs respectively, see figure 3 and below).

The nucleotide sequence information provided herein was obtained by sequencing the WSBV genome using a shotgun sequencing method known in the art. WSBV genomic DNA was initially obtained for sequencing by extraction and purification of viral DNA from infected shrimp tissues using the method of Yang et al. (*J. Virological Methods*, 67:1-4 (1997)), which is hereby incorporated by reference. The nucleotide sequences provided herein are highly accurate, although not necessarily a 100% perfect, representation of the nucleotide sequence of the WSBV genome.

Using the information provided herein together with routine cloning and sequencing methods, one of ordinary skill in the art is able to identify, clone and sequence all "representative fragments" of interest including open reading frames (ORFs) encoding a large variety of WSBV proteins. In very rare instances, this may reveal a nucleotide sequence error present in the nucleotide sequence disclosed herein. Thus, once the present invention is made available (i.e., the information in the Sequence Listing and figures in a useable form), resolving a rare sequencing error would be well within the skill of the art. Nucleotide sequence editing software is publicly available.

Even if all of the very rare sequencing errors in the sequences herein disclosed were corrected, the resulting nucleotide sequence would still be at least 90% identical, and more likely 99% identical, and most likely 99.99% identical to the nucleotide sequence provided herein.

Thus, the present invention further provides nucleotide sequences that are at least 90% identical, or greater, to the nucleotide sequences of the present invention in a form which can be readily used, analyzed and interpreted by the skilled artisan. Methods for determining whether a nucleotide sequence is at least 90% identical to the nucleotide sequence of the present invention are routine and readily available to the skilled artisan. For example, the well known BLAST algorithm can be used to generate the percent identity of nucleotide sequences.

The present invention further provides a prediction of all of the genes within the WSBV genome. This information is provided in Figure 3. The information in the figures can be used to generate WSBV detection kits, antiviral agents, expression arrays, microfluidic devices, individual gene fragments, proteins, antibodies, promoters, protein and nucleotide based viral screens and the like, and to identify commercially important genes and gene products.

Specific Embodiments

Computer Related Embodiments

The nucleotide sequences provided in the present invention, a representative fragment thereof, or nucleotide sequences at least 90% identical to these sequences, may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, that contains a nucleotide sequence of the present invention, i.e., the nucleotide sequences provided in the present invention, a representative fragment thereof, or nucleotide sequences at least 90% identical to these sequences. Such a manufacture provides the WSBV genome or a subset thereof (e.g., a WSBV open reading frame (ORF)) in a form that allows a skilled artisan to examine the manufacture using means not directly applicable to examining the WSBV genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. One such medium is provided with the present application, namely, the present application contains computer readable medium (CD-R) that has the sequence contigs provided/recorded thereon in ASCII text format in a Sequence Listing.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for

recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as OB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide sequences of the present invention, a representative fragment thereof, or nucleotide sequences at least 90% identical to these sequences, in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Software which implements the BLAST (Altschul *et at, J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et at, Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system an be used to identify open reading frames (ORFs) within the WSBV genome that contain homology to ORFs or proteins from other organisms. Such ORFs are protein-encoding fragments within the WSBV genome and are useful in producing commercially important proteins such as proteins used as drug or antiviral targets.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the WSBV genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. Such system

can be changed into a system of the present invention by utilizing the sequence information provided on the CD-R, or a subset thereof without any experimentation.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs that are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the WSBV genome which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are available and can be used in the computer-based systems of the present invention. Examples of such software include, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments of the WSBV genome, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) is chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the WSBV genome possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the WSBV genome. Software which implements the BLAST and BLAZE algorithms (Altschul et al., J Mol. Biol. 215:403-410 (1990)) can be used to identify open reading frames within the WSBV genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention.

One application of this embodiment is provided in Figure 1. Figure 1 provides a block diagram of a computer system 102 that can be used to implement the present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD- ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable storage medium 116 once inserted in the removable medium storage device 114.

The nucleotide sequences of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. Software for accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108 during execution.

Biochemical Embodiments

Nucleic Acid Fragments

Another embodiment of the present invention is directed to isolated fragments of the WSBV genome. The fragments of the WSBV genome of the present invention include, but are not limited to, fragments that encode peptides, hereinafter open reading frames (ORFs) and fragments which modulate the expression of an operably linked ORF. Some of these fragments are identified and described in Figure 3. The isolated nucleic acid molecules of the present invention include, but are not limited to, single stranded and double stranded DNA, and single stranded RNA.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the WSBV genome or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule.

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the WSBV genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to, methods that separate constituents of a solution based on charge, solubility, or size.

In one embodiment, WSBV DNA can be mechanically sheared to produce fragments of about 2kb, 10kb, or 15-20 kb in length. These fragments can then be used to generate a WSBV library by inserting them into vectors, such as plasmid or lambda vectors, using methods well known in the art. Primers flanking each fragment, for example an ORF, can then be generated using nucleotide sequence information provided in the present invention. PCR cloning can then be used to isolate the ORF from the WSBV DNA library. PCR cloning is well known in the art. Thus, given the availability of the present identified gene coding sequences of the WSBV genome, it is routine experimentation to isolate any ORF, or other fragment of the assembly of the present invention, particularly using the information provided in Figure 3. Such fragments can be applied to an array, microfluidic device, or other detection kit format and used to detect expression of a viral gene (see below).

As used herein, an "open reading frame" (ORF) means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein. A

skilled artisan can readily identify ORFs in the WSBV genome using the gene coding sequences provided herein and/or the computer-based systems of the present invention.

As used herein, an "expression modulating fragment" (EMF) means a series of nucleotide molecules which modulates the expression of an operably linked ORF or another EMF.

As used herein, a viral sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of viral EMFs are fragments which induce the expression of an operably linked viral ORF in response to a specific host regulatory factor or physiological event, such as a host anti-viral response.

EMF sequences can be identified within the WSBV genome by their proximity to the ORFs identified using the computer-based systems of the present invention. EMFs may be found immediately 5' to the ORF. Alternatively, EMFs can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site 5' to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. An EMF will modulate the expression of an operably linked marker sequence. A sequence that is suspected of being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include variations thereof. Variations can be routinely determined by comparing the sequence provided in the present invention, or a representative fragment thereof, with a sequence from another WSBV isolate. Furthermore, to accommodate the degeneracy of the genetic code, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another that encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands). Alternatively, error screening, or variant detection, can be performed by sequencing corresponding polynucleotides of WSBV origin isolated by using part or all of the fragments in question as a probe or primer.

Nucleic Acid Fragment Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as hybridization probes for viral messenger RNA, viral transcript/cDNA, and viral genomic DNA to isolate full-length viral cDNA and viral genomic clones encoding the peptides described in Figure 3, and for use in viral screens and antiviral agents.

Oligonucleotide probes have long been used to detect complementary nucleic acid sequences in a nucleic acid of interest (the "target" nucleic acid) in the form of detection kits/reagents. In some assay formats, the oligonucleotide probe is tethered, i.e., by covalent attachment, to a solid support, and arrays of oligonucleotide probes immobilized on solid supports have been used to detect specific nucleic acid sequences in a target nucleic acid. See, e.g., PCT patent publication Nos. WO 89/10977 and 89/11548. In other formats, the detection reagents are supplied in solution.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the figures. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

Each of the ORFs of the WSBV genome that can be routinely identified using the computer system of the present invention can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes or diagnostic amplification primers to detect the expression of a particular gene or groups of genes. This is particularly useful in the form of nucleic acid arrays employing 1 or more, 10 or more, 100 or more, or most to all of the WSBV ORFs in a single array.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful for synthesizing antisense molecules of desired length and sequence.

The nucleic acid molecules are useful as primers for the 5' nuclease PCR assay (hereafter referred to as the TaqMan assay). The TaqMan assay provides a sensitive and rapid means of detecting viral nucleic acid and therefore is well suited for use in viral screening

applications such as detection kits. The TaqMan assay detects the accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye at the 5' end of the probe and a quencher dye at the 3' end of the probe. During the PCR reaction, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The 5' nuclease activity of DNA polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target and is amplified during PCR. Therefore, only the target sequence of interest is detected.

Preferred TaqMan primer and probe sequences are disclosed in Figure 3. It will be apparent to one of skill in the art that the disclosed primers and probes of the present invention are useful as diagnostic probes or amplification primers for screening for the presence of WSBV in a biological sample or for isolating or screening particular WSBV genes.

The nucleic acid molecules are also useful for expressing antigenic portions of the WSBV proteins that can then be used, for example, to develop antibodies to the viral antigens.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of WSBV nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific viral nucleic acid molecule, either DNA or RNA, in cells and tissues of shrimp or other organisms under moderate or stringent conditions. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 °C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65 °C. Examples of moderate to low stringency hybridization conditions are well known in the art. Furthermore, probes corresponding to the viral peptides described herein can be used to assess expression and/or gene copy number in a given infected cell, tissue, or organism. For example, Northern blots can be used for RNA detection, Southern blots can be used for DNA detection, and Western blots can be used for peptide/protein detection. These uses are relevant for detecting the presence of virus in shrimp as well as for monitoring the distribution of virus throughout various cells and tissues of shrimp during the course of viral infection.

In addition, each of the expression modulating fragments (EMFs) can be used in DNA-protein binding assays to screen for modulating peptides which may be present in cells and

tissues of shrimp or other organisms. This use is relevant for obtaining the specific host regulatory factors that interact with promoters in the WSBV genome.

Nucleic acid expression assays are also useful for drug screening to identify compounds that modulate viral nucleic acid expression. The invention thus provides a method for identifying a compound that can be used to treat viral infection. The method typically includes assaying the ability of the compound to modulate the expression of viral nucleic acid and thus identifying a compound that can be used to treat a viral infection. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells infected with virus particles or recombinant cells genetically engineered to express specific viral nucleic acid sequences. Cell-free assays can be used to detect the ability of a compound to directly bind to a nucleic acid fragment or protein.

The assay for viral nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Furthermore, the expression of host cell genes that are up- or down-regulated in response to the viral protein can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of viral gene expression can be identified by a method wherein a cell infected with virus is contacted with a candidate compound and the expression of viral mRNA determined. The level of expression of viral mRNA in the presence of the candidate compound is compared to the level of expression of viral mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of viral nucleic acid expression based on this comparison and be used, for example, to disrupt viral replication. When expression of viral mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of viral nucleic acid expression. When viral nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of viral nucleic acid expression.

The invention further provides methods of treating viral infection, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate viral nucleic acid expression in cells and tissues infected with the virus. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) of nucleic acid expression. Generally, viral nucleic acid expression is down-regulated to prevent viral replication and treat viral infection.

Alternatively, a modulator for viral nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule modulates viral nucleic acid expression in the cells and tissues infected with the virus.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the viral gene in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a virus can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the virus to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the virus has not become resistant.

The nucleic acid molecules are also useful as antisense constructs to control viral gene expression in infected cells and tissues. A DNA antisense nucleic acid molecule is designed to be complementary to, and therefore bind to, a region of the viral gene necessary for transcription, thereby preventing transcription and hence production of viral protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the viral mRNA and thus block translation of viral mRNA into protein by the host cell's translational machinery. Alternatively, a class of antisense molecules can be used to inactivate viral mRNA in order to decrease expression of viral nucleic acid and inhibit viral replication or function. These molecules can therefore be used to treat viral infection. This technique involves cleavage of viral mRNA by ribozymes that recognize one or more regions of viral mRNA that attenuate the ability of the mRNA to be translated by host cell translational machinery. Possible regions include coding or control regions; particularly coding or control regions encoding or regulating proteins that play critical roles in viral function or replication, such as entry into the nucleus of the host cell or virion assembly.

The nucleic acid molecules of the present invention can be employed to create transgenic viral resistant shrimp. Several possible mechanisms could be employed to impart WSBV resistance to shrimp using the nucleic acid and protein coding sequences provided by the present invention. One possible mechanism of imparting WSBV resistance to shrimp involves transforming shrimp cells with viral nucleic acids that express an attenuated virion coat protein such that when the transgenic shrimp is infected with WSBV, the expressed coat protein envelopes the virus and thereby prevents translation of the viral DNA. In this example, the virion coat protein can either be constitutively expressed or regulated by a promoter that is activated upon WSBV infection. Shrimp cells can be transformed with viral DNA under

suitable conditions known in the art. The WSBV construct in a vector can be microinjected directly into host cells using micropipettes, [Crossway, *Mol. Gen. Genetics*, 202:179-85 (1985)], or using polyethylene glycol [Krens *et al., Nature*, 296:72-74 (1982)]. Alternatively, shrimp cells may be transformed by incubating the shrimp cells or tissue with an inoculum of bacteria that have been transformed with a vector comprising a gene that imparts WSBV resistance. The transformed shrimp cells are then grown and regenerated into shrimp such that the proteins expressed by the transformed cells impart WSBV resistance to the shrimp.

Nucleic Acid Arrays and Detection Reagents

The present invention further provides detection reagents and kits, such as arrays or microarrays, of nucleic acid molecules that are based on the novel WSBV sequence information provided in the present invention and particularly the transcript information (SEQ ID NOS: 2, 4, 6. . . 280, 282, 284, 286-293) provided in Figure 3 and polymorphic sites on these transcripts summarized in Table 1.

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid, or semi-solid support. The development of arraying technologies such as photolithographic synthesis of a nucleic acid array and high density spotting of cDNA products has provided methods for making very large arrays of oligonucleotide probes in very small areas. See U.S. Pat. No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. Microfabricated arrays of large numbers of oligonucleotide probes, called "DNA chips", offer great promise for a wide variety of applications. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. *et al.* (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be

preferable to use oligonucleotides that are only 7-20 nucleotides in length. For others, such as cDNA, longer lengths are possible and preferable. These can be of the order of 1kb or more.

The microarray or detection kit may contain oligonucleotides that cover the known 5' or 3' sequence, sequential oligonucleotides that cover the full-length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a viral gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the viral gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from one to two millions. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedure. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

In other embodiments, the array or detection reagent/kit can be produced by spotting a nucleic acid molecule onto the surface of a substrate (See Brown et. al., US Patent No.

5,807,522). In such embodiments, PCR primers to one or more nucleic acid fragments are used to generate nucleic acid molecules suitable for deposition onto a substrate.

In order to conduct sample analysis using a microarray or detection kit, viral nucleic acid is isolated from a biological sample infected with WSBV and the viral nucleic acid is made into hybridization probes. Viral nucleic acid may be isolated from biological samples obtained from fluids, cultured cells, biopsies, or other tissue preparations from a shrimp or other organism of interest that is infected with WSBV. Viral mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct viral sequences simultaneously. This data may be used for large scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among viral isolates.

Using such arrays, the present invention provides methods to identify the expression of one or more of the ORFs of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising most, if not all of the genes in the WSBV genome, or rationally selected subsets thereof. The genomic sequence (SEQ ID NO: 1) and transcript sequences (SEQ ID NOS: 2, 4, 6. . .280, 282, 284, 286-293) of the WSBV genome of the present invention are provided in Figure 2 and Figure 3 and polymorphic sites on these transcripts summarized in Table 1.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the WSBV genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to*

Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include, but are not limited to, nucleic acid extracts, cells, and protein or membrane extracts of cells infected with WSBV. The test sample used in the above-described method will vary based on the assay format, the nature of the detection method, and the tissues, cells, or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or for preparing cells are well known in the art and can readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the WSBV genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting the presence of a bound nucleic acid. Preferred kits will include detection reagents/arrays/chips/microfluidic devices that are capable of detecting the expression of 1 or more, 10 or more, 100 or more, or most or all of the genes expressed in WSBV, particularly the genes provided in Figure 3.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers may include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified ORFs that can be routinely identified using the sequence information disclosed herein can be

readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Protein/Peptide Molecules

The present invention provides nucleic acid sequences that encode WSBV protein molecules (Figure 3). The peptide sequences provided in Figure 3, as well as the obvious variants described herein, and using the information in Figure 3, will be referred to herein as the WSBV peptides of the present invention or peptides/polypeptides/proteins of the present invention.

Enzymes and other viral proteins are produced during viral activity and replication and can be used as targets for screening and quantitating a particular virus, or as antiviral targets. Like viruses in general, WSBV utilizes the resources of the host cell for production of viral proteins. Viral proteins can be detected using an antibody, or binding portion thereof, to the protein or a probe that recognizes proteins or peptides of the present invention. Viral antigens present either on the surface or within the infected cell can be detected by various antibody tests, including immunoflourescence or enzyme immunoassay (EIA). Virus or antigen released from infected cells can be detected by such antibody tests as enzyme linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or latex agglutination (LA). Protein-based tests such as these for WSBV antigens are useful for detecting outbreaks of WSBV in shrimp populations.

The present invention provides isolated peptide and protein molecules that comprise, consist essentially of, or consist of the amino acid sequences of the WSBV peptides disclosed in Figure 3, (which are encoded by the transcript sequences that are also shown in Figure 3), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components.

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other

proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a WSBV peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

An isolated WSBV peptide can be purified from cells infected with WSBV, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule encoding the peptide can be cloned into an expression vector, the expression vector introduced into a host cell, and the protein expressed in the host cell. The protein can then be isolated from the host cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 3 (SEQ ID NOS:3, 5, 7...281, 283, 285), for example, proteins encoded by the transcript/cDNA nucleic acid sequences also shown in Figure 3 (SEQ ID NOS:2, 4, 6...280, 282, 284, 286-293). A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 3 (SEQ ID NOS:3, 5, 7...281, 283, 285), for example, proteins encoded by the transcript/cDNA nucleic acid sequences also shown in Figure 3 (SEQ ID NOS:2, 4, 6...280, 282, 284, 286-293). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 3 (SEQ ID NOS:3, 5, 7...281, 283, 285), for example, proteins encoded by the transcript/cDNA nucleic acid sequences also shown in Figure 3 (SEQ ID NOS:2, 4, 6...280, 282, 284, 286-293). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the

protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a WSBV peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the WSBV peptide. "Operatively linked" indicates that the WSBV peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the WSBV peptide.

In some uses, the fusion protein does not affect the activity of the WSBV peptide per se. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant WSBV peptide. In certain host cells, expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel et al., *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A WSBV peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the WSBV peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such

variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the WSBV peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog protein family and the evolutionary distance between orthologous viruses.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In preferred embodiments, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which are introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14,

12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res.* 12(1):387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other viruses related to WSBV or functionally related protein sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. The results of one such analysis are provided in Figure 3.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the WSBV peptides of the present invention as well as being encoded by the same viral gene as the WSBV peptide provided herein.

Variants of a WSBV peptide can readily be identified as being a WSBV protein having a high degree of sequence homology/identity (also referred to as "significant sequence homology") to at least a portion of the WSBV peptide as well as being encoded by the same viral gene as the WSBV peptide provided herein. Viral genes can readily be determined based on the WSBV sequence information provided in Figure 3. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present

invention, will be encoded by a nucleic acid sequence that will hybridize to a WSBV peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Paralogs of a WSBV peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the WSBV peptide, as being encoded by a gene from WSBV, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences typically share at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a WSBV peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a WSBV peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the WSBV peptide as well as being encoded by a gene from another virus. Preferred orthologs will be isolated from viruses of commercial or medical importance for the development of broad-spectrum diagnostic and anti-viral agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a WSBV peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two viruses yielding the proteins.

Non-naturally occurring variants of the WSBV peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to, deletions, insertions, and substitutions in the amino acid sequence of the WSBV peptide. For example, one class of substitutions is conserved amino acid substitutions. Such substitutions are those that substitute a given amino acid in a WSBV peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al., Science* 247:1306-1310 (1990).

Variant WSBV peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind to host cell receptors or ability to form structural components such as the viral nucleocapsid or outer membrane, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 3 provides the results of protein analysis and can be used to identify critical domains/regions.

Functional variants can also contain substitutions of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 3. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as DNA binding. Sites that are critical for virus/host cell receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al. Science* 255:306-312 (1992)).

The present invention further provides fragments of the WSBV peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 3. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a WSBV peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the WSBV peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the WSBV peptide, e.g., active site or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis).

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes during the course of viral infection, such as processing and other post-translational modifications by the host cell, or by chemical

modification techniques well known in the art. Common modifications that occur naturally are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Examples of known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan et al. (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Peptides or protein encoding sequences of the present invention can be modified or mutated, either naturally, such as by host cell mechanisms, or by techniques known to those of skill in the art, to disrupt protein formation or protein function, and thereby disrupt viral replication and function. These methods can be used to prevent and/or treat viral infection.

Accordingly, the WSBV peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature WSBV peptide is fused with another compound, such as a compound to increase or decrease the half-life of the WSBV peptide (for example, polyethylene glycol), in which the additional amino acids are fused to the mature WSBV peptide, such as a leader or secretory sequence or a sequence for purification of the mature WSBV peptide or a pro-protein sequence, or in which the WSBV peptide has been modified or mutated, either naturally or recombinantly, to disrupt protein function, and thereby disrupt WSBV function and/or replication.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological samples; and as markers for infected samples in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of viral infection). Where the protein binds or potentially binds to another protein or ligand (such as, for example, a host cell receptor protein), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the viral peptides of the present invention are based primarily on the function of the protein. For example, isolated WSBV peptides serve as targets for identifying antiviral agents, particularly for identifying antiviral agents that interfere with viral replication in a host cell infected with a virus that expresses the peptide. Specific and substantial uses for the molecules of the present invention are provided herein. Further uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays for viruses that are related to WSBV. Such assays involve any of the known protein functions or activities or properties useful for diagnosis of WSBV infection.

The proteins of the present invention are also useful in virus screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., host cells infected with the virus, as a biopsy or expanded in cell culture. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the viral protein. Cell-based or cell-free systems can be used in assays for protein activity, such as enzymatic activity. Cell-free assays can be used to

detect the ability of a compound to directly bind to a protein or nucleic acid fragment of the present invention.

The polypeptides can be used to identify compounds that modulate activity of the protein. Both the WSBV peptides of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the WSBV peptide. These compounds can be further screened against a functional WSBV peptide to determine the effect of the compound on the WSBV peptide activity. Further, these compounds can be tested in shrimp to determine activity/effectiveness. Compounds can be identified that inactivate the WSBV peptide to a desired degree (antagonists).

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the WSBV protein and a target molecule that normally interacts with the WSBV protein, e.g. a host cell receptor. Such assays typically include the steps of combining the WSBV protein with a candidate compound under conditions that allow the WSBV protein, or fragment thereof, to interact with the target molecule, and detecting the formation of a complex between the WSBV protein and the target or detecting the biochemical consequence of the interaction between the WSBV protein and the target, such as any of the associated effects of host cell signal transduction such as protein phosphorylation, cAMP turnover, or adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a non-virulent soluble fragment of the WSBV peptide that competes for substrate binding, such as for binding to shrimp cellular receptors. Other candidate compounds include non-virulent mutant WSBV peptides or appropriate fragments containing mutations that prevent WSBV virulence and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but is inactive or non-virulent, is encompassed by the present invention.

The invention further includes other end point assays to identify compounds that inhibit WSBV activity. The assays typically involve an assay of events in the shrimp cell signal transduction pathway that indicate viral activity. Thus, the phosphorylation of a substrate, activation of a protein, a change in the expression of genes that are up- or down-regulated in a host cell in response to WSBV infection can be assayed.

Any of the viral functions mediated by a WSBV protein can be used as an endpoint assay. These include all of the biochemical or biological events described herein, and in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the figures, particularly Figure 3.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the viral protein (e.g. binding partners and/or ligands). Thus, a compound is exposed to a viral polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble viral polypeptide is also added to the mixture. If the test compound interacts with the soluble viral polypeptide, it decreases the amount of complex formed or activity from the viral protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the viral protein. Thus, the soluble polypeptide that competes with the target viral protein region is designed to contain peptide sequences corresponding to the region of interest. See Hodgson, *Bio/technology*, 1992, Sept 10(9), 973-80, for a review of competition binding assays and other receptor screening assays.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the viral protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing viral proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the viral protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and a candidate drug compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the

complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of a viral protein target ligand, such as a host cell receptor protein, found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the viral protein or its target ligand can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the viral protein but which do not interfere with binding of the viral protein to its target ligand can be derivatized to the wells of the plate, and the viral protein trapped in the wells by antibody conjugation. Preparations of a viral protein target ligand and a candidate compound are incubated in the viral protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the viral protein target ligand, or which are reactive with viral protein and compete with the target ligand, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target ligand.

Agents that modulate one of the viral proteins of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in a shrimp, or other organism, infected with WSBV.

Modulators of viral protein activity identified according to these drug screening assays can be used to treat shrimp infected with WSBV. These methods of treatment include the steps of administering a modulator of viral protein activity in a pharmaceutical composition to an organism, such as a shrimp, that is infected with WSBV, the modulator being identified as described herein.

In yet another aspect of the invention, the WSBV proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins which bind to or interact with the viral protein and are involved in viral protein activity, and therefore are targets for inhibiting viral protein activity.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a viral protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g.,

GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a viral protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the viral protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model, such as a shrimp infected with WSBV. For example, an agent identified as described herein (e.g., a viral protein-modulating agent, an antisense viral nucleic acid molecule, a viral protein-specific antibody, or a viral protein-binding partner) can be used in a shrimp, or other organism, infected with WSBV to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The viral proteins of the present invention are also useful for providing targets for diagnosing viral infection. Accordingly, the invention provides methods for detecting the presence, or levels of, the viral protein (or encoding nucleic acid) in an infected cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the viral protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to a WSBV protein. A biological sample includes tissues, cells and biological fluids isolated from a shrimp or other infected organism, as well as tissues, cells and fluids present within the infected organism.

In vitro techniques for detection of viral peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can

be detected in vivo in an infected organism by introducing into the subject a labeled anti-peptide antibody or other type of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in an infected organism can be detected by standard imaging techniques. Particularly useful are methods that detect fragments of a peptide in a sample.

Antibodies

The invention also provides antibodies that selectively bind to one of the WSBV peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989). In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The antibodies generated by the organism in response to the immunogen are then isolated. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 3, and domains of sequence homology or divergence between WSBV and other viruses, such as those that can readily be identified using protein alignment methods and as presented in the figures.

Monoclonal antibodies can be produced by hybridomas, which are immortalized cell lines capable of secreting a specific monoclonal antibody. The immortalized cell lines can be created in vitro by fusing two different cell types, usually lymphocytes, one of which is a tumor cell.

Antibodies are preferably prepared from regions or discrete fragments of the WSBV proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or virus/host interaction. Figure 3 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments corresponding to regions that are located on the surface of the protein, e.g., hydrophilic regions, or can be selected based on sequence uniqueness (see Figure 3).

Detection of an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural viral protein from host cells infected with WSBV, as well as recombinantly produced protein. In addition, such antibodies are useful for detecting the presence of the viral proteins of the present invention in cells or tissues in order to determine the pattern of viral infection among various cells or tissues in a shrimp or other organism over the course of viral infection. Further, such antibodies can be used to detect protein in situ, in vitro, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of viral infection.

The antibodies can also be used to assess subcellular localization of virus particles in host cells. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at preventing or halting expression of the WSBV protein, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

The antibodies are also useful diagnostic tools, such as for use as immunological markers for aberrant viral protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for inhibiting protein function. For example, antibodies may bind directly to viral peptides to block binding of the viral peptide to a binding partner such as a host cell receptor. Antibodies can thereby serve as antiviral agents. Antibodies can be prepared against specific fragments containing sites required for protein function or against intact viral protein that is associated with virulence.

The invention also encompasses kits for using antibodies to detect the presence of a WSBV protein in a biological sample, such as a shrimp cell sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting viral protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail above for nucleic acid arrays and similar methods have been developed for antibody arrays.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claim.

Certain aspects of the present invention are described in greater detail in the non-limiting examples that follow.

Examples:

Infected prawn

Dead and moribund *P. japonicus* with evident white spots on the inside surface of the crust were collected from a prawn farm and kept at 4 ^{0}C.

Isolation of nucleocapsids

Hepatopancreata gill and intestine were removed from *P.japonicus* and placed in an ice-bathed beaker, homogenized as a 10% suspension in TESP buffer (50 mmol/l, Tris-HCl, pH 8.5, 10 PMSF), then centrifuged at 6500 x *g* for 10 min at 4 ^{0}C. The supernatant was recentrifuged at 30,000 x *g* for 30 min at 4^{0}C. The pellet was suspended in an approximate two volume of TESP buffer containing 1% (v/v) Triton X-100. After centrifugation at 5000 x *g* for 10 min, the supernatant was centrifuged again at 25,000 x *g* for 20 min. The pellet was suspended in TESP buffer and differential centrifugation was repeated, and then the precipitate was resuspended in TMP (100 mmol/l, Tris-HCl, pH 7.5, 10 mmol/l $MgCl_2$/l, 1 mmol/l PMSF) buffer and treated with DNase and RNase. The mixture was incubated at 37^{0}C for 15 min and 30 ml of TESP buffer was added. Differential centrifugation was repeated again and the pellet was resuspended in 1 ml of TESP buffer, 1 μl of suspension was dropped on a copper grid, negatively stained with 2% (w/v) uranyl acetate, pH 7.6, and observed using a JEM-100CX II transmission electron microscope.

Purification of viral DNA

The nucleocapsid suspension was lysed with 2 ml of GTE buffer (6 mol/l guanidine hydrochloride; 50 mmol/l Tris-HCl; 10 mmol/l EDTA; pH 7.0), slightly mixed, and then centrifuged at 25,000 x *g* for 10 min at 4 ^{0}C. The supernatant was collected and 0.02 vol. of 1 mol/l $MgCl_2$ and 0.6 vol. of isopropanol were added. After centrifugation, the pellet was picked out with pipette tip, washed twice with 70% ethanol and then dissolved in 1 ml of TE buffer containing 0.5% (w/v) SDS and 0.5 mg/ml proteinase K and incubated at 55^{0}C for 2-3 hrs. The DNA was precipitated again with 0.01 vol. of 1 mol/l $MgCl_2$ and 0.25 vol. of isopropanol and dissolved in 0.1 x TE buffer. The viral DNA obtained was quantified by a spectrophotometer.

WSBV genomic DNA library construction

Construction of a random "shotgun" library:

WSBV genomic DNA was sheared with sonication. Mung Bean nuclease was used to blunt the end. The DNA fragments between 1.8-2kb were recovered from an agarose gel following electrophoresis. The blunt end DNA was cloned into pUC18 vectors. The vector was subsequently transformed into DH5α cells and plated onto LB plate.

Construction of a restriction fragment library:

WSBV genomic DNA was partially digested with Sau3A1 restriction enzymes. DNA fragments between 5-10kb were recovered from the agarose gel. pBluescript vectors were digested by the restriction enzyme and the ends were dephosphorylated. The fragments were cloned into pBluescript vector and transformed into XL-blue competent cells. Subsequently, the DNA plasmid was prepared.

Large scale DNA sequencing

PCR reactions:

PCR reactions were carried out in a 25 μl volume containing 0.2mM dNTP, 1.5mM $MgCl_2$, 5 μM of each primers, 2.5 unit of Taq polymerase, and a single white colony as template. Take out 1μl as glycerol stocks. PCR reactions were done in a PE 9700. The cycling profile consisted of an initial denaturation at 95°C for 12 min (one cycle) followed by 30 cycles of denaturation at 95°C for 15s, annealing at 58°C for 20s, and extension at 72°C for 2 min. Then 5 μl of PCR products were visualized on 1% agarose gels stained with ethidium bromide. Excess primers and dNTPs were removed by digesting the PCR products with exonuclease I and shrimp alkaline phosphatase.

DNA sequencing:

Sequencing reactions were carried out in 5µl volume containing 1µl of "BigDye" premix (PE Applied Biosystems), 3.2µM sequencing primer and 30-90ng PCR products. Sequencing reactions were done in a PE 9700. The cycling profile consisted of 30 cycles of denaturation at 95°C for 30s, annealing at 50°C for 30s, and extension at 60°C for 4 min. Excess Dye terminators were remove with ethanol precipitation, and sequencing was carried out on ABI 377 automated sequencer.

Sequence analysis

A total of 5795 sequences were assembled in a UNIX system using InnerPeace software designed based on the "Phred, Phrap and Consed" program originally developed by the University of Washington. Sequences were edited and finished as follows: a) for bad

sequence quality, sequencing was repeated; b) for regions with repetitive sequences, which may cause misassembly, primers were designed for walking on the original PCR products or on plasmid DNA; c) for mapping gaps, clones were sequenced that cover the gap; d) for physical gaps, PCR primers were designed between the gaps, then the PCR products were sequenced that cover the gap; e) for gaps that can't be covered by PCR methods, walking on WSBV genomic DNA was applied.

WSBV cDNA library construction and WSBV cDNA clone selection

Poly(A)-mRNA was purified by using the "PolyATtract System1000" kit (Promega). Double stranded cDNA was synthesized and cloned using the "SUPERSCRIPT™ Plasmid System for cDNA Synthesis and Plasmid Cloning" kit (GIBCO BRL). cDNA clones were transformed into DH10□ cells and then plated. WSBV cDNA clones were selected by DNA hybridization using Dig labeled WSBV genomic DNA as a probe (Dig labeling kit, Boehringer Mannheim). Finally, the plasmid DNA was prepared for automatic sequencing.